SEMIARID PRECIPITATION FREQUENCY STUDY

Update of *Technical Paper No. 40, Technical Paper No. 49* and *NOAA Atlas 2*

Twentieth Progress Report
1 January 2002 through 31 March 2002

Hydrometeorological Design Studies Center
Hydrology Laboratory

Office of Hydrologic Development
U.S. National Weather Service
Silver Spring, Maryland

April 2002

DISCLAIMER

The data and information presented in this report should be considered as preliminary and are provided only to demonstrate current progress on the various technical tasks associated with this project. Values presented herein are NOT intended for any other use beyond the scope of this progress report. Anyone using any data or information presented in this report for any purpose other than for what it was intended does so at their own risk

TABLE OF CONTENTS

SEMIARID PRECIPITATION FREQUENCY STUDY

Update of *Technical Paper No. 40, Technical Paper No. 49* and *NOAA Atlas 2*

1. Introduction.

The Hydrometeorological Design Studies Center (HDSC), Hydrology Laboratory, Office of Hydrologic Development, U.S. National Weather Service is updating its precipitation frequency estimates for the Semiarid Southwestern United States. Current precipitation frequency estimates for the Semiarid region are contained in *Technical Paper No. 40* "Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years" (Hershfield 1961), *Technical Paper No. 49* "Two- to ten-day precipitation for return periods of 2 to 100 years in the contiguous United States" (Miller et al 1964) and *NOAA Atlas 2* “Precipitation-Frequency Atlas of the Western United States.” The new study includes collecting data and performing quality control, compiling and formatting datasets for analyses, selecting applicable frequency distributions and fitting techniques, analyzing data, mapping and preparing reports and other documentation.

The study will determine annual and seasonal precipitation frequencies for durations from 5 minutes to 60 days, for return periods from 2 to 1000 years. The study will review and process all available rainfall data for the Semiarid study area and use accepted statistical methods. In particular, the Semiarid Study is the pilot study in which decisions regarding the methods and format are being made that will affect subsequent studies. The study results will be published as Volumes of *NOAA Atlas 14*. They will also be made available on the internet using web pages with the additional ability to download digital files.

The Semiarid study will produce estimates for 4 states completely, Arizona, Nevada, New Mexico, and Utah, and southeastern California. Additional data from 7 bordering states and Mexico (Figure 1) are included for continuity across state borders. The core and border areas, as well as original regions proposed for this analysis, are shown in Figure 1. Many of these original regions have been subdivided because they were found to be heterogeneous. Warm and Cool season months are also shown in boxes in each region in Figure 1.

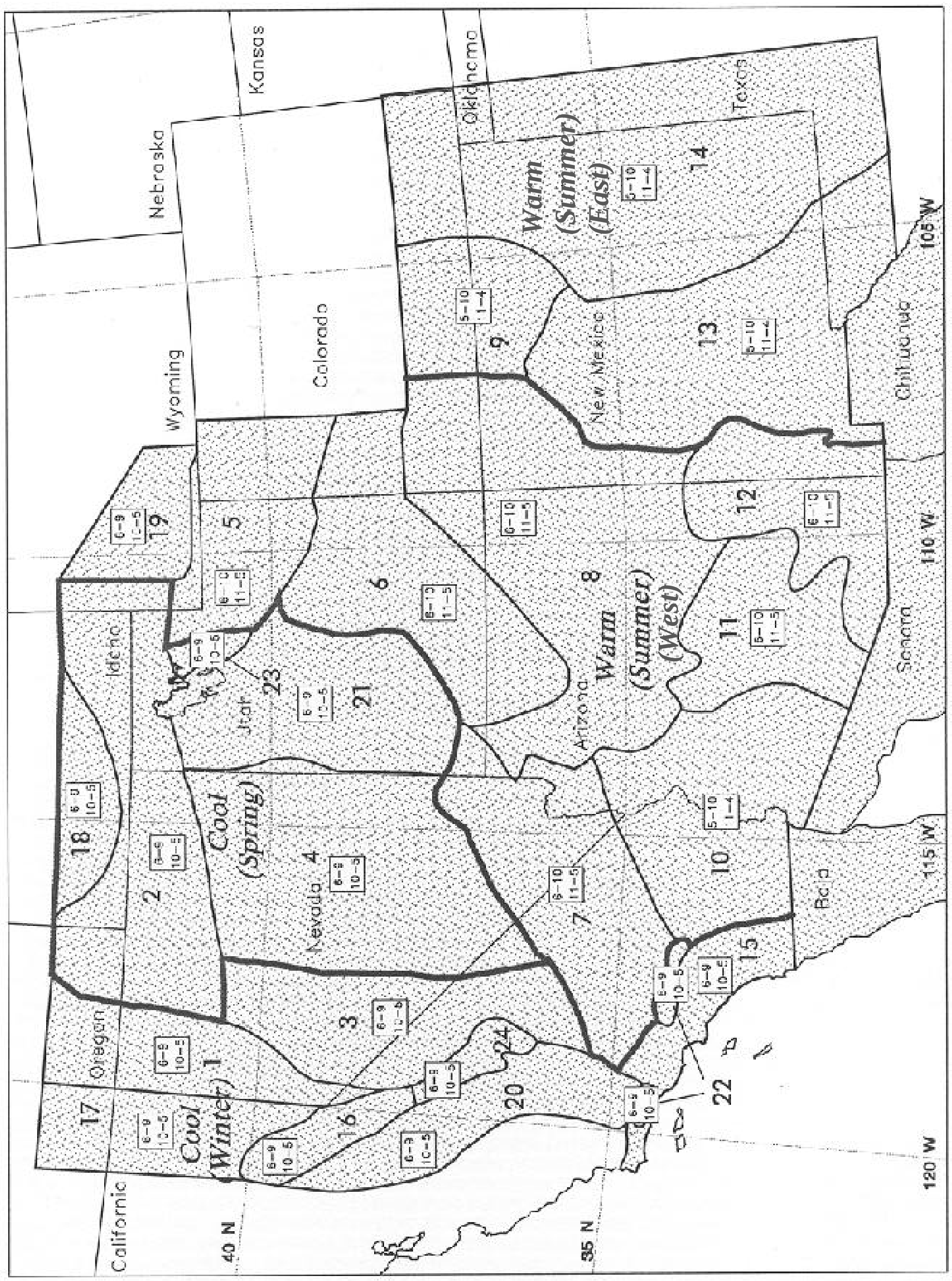

Figure 1. Semiarid Precipitation Frequency study area and original region boundaries.

2. Highlights.

The quality control of the updated daily dataset (primarily 1949 - 2000) is complete. With this update, we benefitted from the improved quality of daily data from the National Climatic Data Center and the application of our current QC methods. Annual maximum series for the 1-day (24-hour) event were extracted and analyzed. We have developed and tested software to extract annual maximums of longer duration events (4-day through 60-day). The 1-day annual maximum series of all data records were screened for large gaps in time using specified "Gap Check" criteria. Stations were merged and adjustments were made to produce more congruent data records. Additional information is provided in Section 4.1, Data Collection and Quality Control.

L-moment analysis is complete for n-minute and hourly data. Initial L-moment analyses have been conducted on 1-day annual maximum series for all 24 original regions in the Semiarid study. Most regions were found to be heterogeneous. We are currently subdividing the regions to form smaller homogeneous regions. Cases where observed maximums at stations exceed 100-yr return frequency estimates are being investigated during this process as well. Additional information is provided in Section 4.2, L-moment Analysis and Regionalization.

Cross-boundary differences in the statistics, notably the Regional Growth Factor (RGF), of the subdivisions of regions 10, 11, and 12 are still rather large. Further subdivision was employed to mitigate this issue, but the end result may require spatial smoothing to compensate for these differences. Additional information is provided in Section 4.2, L-moment Analysis and Regionalization.

Development is underway to add functionality to the Precipitation Frequency Data Server (PFDS) to extract station-specific data. This functionality will allow for a review of the point-precipitation frequency estimates before the interpolated grids are finalized. Additional information is provided in Section 4.3, Precipitation Frequency Data Server.

Depth Area Duration (DAD) values will be prepared and presented in a separate report. We are currently gathering and formatting data from geographically spaced dense area rain gage networks (DRNs) across the United States. Additional information is provided in Section 4.4, Depth Area Duration Study.

Contract formalities between HDSC and the Spatial Climate Analysis Service (SCAS) at Oregon State University (OSU) have been finalized. As soon as HDSC has finalized the calculations of the point precipitation frequency estimates, efforts will be immediately underway to spatially interpolate with PRISM. Additional information is provided in Section 3.1.3, Mapping Analyses.

3. Status.

3.1 Project Task List.

The following checklist shows the components of each task and an estimate of the percent completed per task. Past status reports should also be referenced for additional information.

Semiarid study checklist [estimated percent complete]:

Data Collection, Formatting and Quality Control [99%]:
- Daily
- Hourly
- 15-minute
- N-minute

L-Moment Analysis/Frequency Distribution for 5 minute to 60 days and 2 to 1000 years [70%]:
- Daily
- Hourly
- 15-minute
- N-minute

Spatial Interpolation [0%]
- Create grids of interpolated means for each duration using PRISM (see Table 1)
- Subject grids of interpolated means to external review
- Create smoothed regional growth factor (RGF) grids using GRASS: (5-1000) yr (1-12) hr, (5-1000) yr 24hr, (5-1000) yr (2-60) day

Table 1. List of Grids of Distributed Means.

Duration	Season
1-hr	all
1-hr	cool, warm
2-hr	all
3-hr	all
6-hr	all
6-hr	cool, warm
12-hr	all
24-hr	all
24-hr	cool, warm
48-hr	all
4-day	all
7-day	all
10-day	all
20-day	all
30-day	all
45-day	all
60-day	all
Total: 26 (14 all, 6 warm, 6 cool)	

Precipitation Frequency Maps [0%]

- Multiply appropriate RGF and distributed mean grids to produce precipitation frequency grids for durations and seasons shown in Table 2
- Apply domain-wide conversion factor to the 1-hour precipitation frequency grids to calculate the n-minute (5-, 10-, 15-, and 30-minute) grids
- Perform internal consistency checks (comparing rasters of sequential duration and frequency)

Table 2. List of Precipitation Frequency Rasters.

Duration	**Frequency**	**Season**
5-min	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
5-min	2-yr, 100-yr	cool, warm
10-min	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
10-min	2-yr, 100-yr	cool, warm
15-min	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
15-min	2-yr, 100-yr	cool, warm
30-min	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
30-min	2-yr, 100-yr	cool, warm
1-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
1-hr	2-yr, 100-yr	cool, warm
2-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
3-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
6-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
6-hr	2-yr, 100-yr	cool, warm
12-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
24-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
24-hr	2-yr, 100-yr	cool, warm
48-hr	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
4-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
7-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
10-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
20-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
30-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
45-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all
60-day	2-yr, 5-yr, 10-yr, 25-yr, 50-yr, 100-yr	all

Data Trend Analysis [10%]

- Analyze linear trends in annual maxima and variance over time
- Analyze shift in means of annual maxima between two time periods (i.e., test the equality of 2 population distribution means)

Seasonal Analysis [75%]

- Create graphs of percentage of precipitation maxima in each month of a year

Temporal Distributions of Extreme Rainfall [100%]

- assemble hourly data by quartile of greatest precipitation amount and convert to cumulative rainfall amounts for each region
- prepare graphs of representative storm-types and seasons

Deliverables [35%]

- Prepare data for web delivery
- Prepare documentation for web delivery
- Write hard copy of Final Report
- Publish hard copy of Final Report

Additional Work:

Spatial Relations (Depth-Area-Duration Study) [20%]

- Obtain data from dense-area reporting networks
- QC and format data from dense networks
- Compute maximum and average annual areal depth for each duration from stations from each network
- Compute ratio of maximum to average depth for all durations and networks and plot
- Draw curves of best fit (depth-area curves) for each duration and network

3.1.1 Data Collection and Quality Control.

The digitized TD3206 daily dataset from NCDC for the time period before 1949 has been added and quality controlled. The bulk of the daily dataset, primarily from 1949 through 2000, has been updated and quality controlled. The hourly and n-minute datasets are also complete.

3.1.2 Precipitation Frequency Analyses.

Annual maximum series were extracted for the n-minute, hourly, and daily data using established criteria. L-moment analysis is complete for the n-minute and hourly data. The daily annual maximum series have been quality controlled for large gaps in records that may skew results. L-moment analysis, including discordancy, heterogeneity, and real data checks, is proceeding with the daily data. Many of the initial Semiarid regions were found to be heterogeneous and are, therefore, being subdivided into smaller homogeneous regions.

3.1.3 Spatial Interpolation.

Contract formalities between HDSC and the Spatial Climate Analysis Service (SCAS) at Oregon State University (OSU) have been finalized. As soon as the point precipitation frequency estimates and mean annual maxima values are calculated by HDSC, SCAS will use PRISM (Parameter-elevation Regressions on Independent Slopes Model) to spatially interpolate the mean annual maxima values (a.k.a. “index flood”) to grids. At HDSC, the “index flood” grids will be multiplied by the appropriate regional growth factor (RGF) grid to derive each of the precipitation frequency grids. We are evaluating different spatial smoothing techniques to mitigate any large RGF boundary differences.

3.1.4 Other Analyses.

Temporal distributions of extreme rainfall in the Semiarid study area have been examined. The results will be presented in graphs representative of different storm-types and seasons. These graphs and accompanying tables are complete.

A new trend analysis will be conducted on the expanded daily dataset. The analysis will include parametric and non-parametric statistical tests to discern if any trend exists between sequential segments of data records. The software for this analysis has been developed.

3.1.5 Precipitation Frequency Data Server.

The Semiarid study station-specific results will be made available on the HDSC Precipitation Frequency Data Server for review. Once mapping is complete and reviewed, the spatially interpolated grids will also be available for review. The Precipitation Frequency Data Server displays precipitation frequency values and intensity-duration-frequency curves and tables. Additional station-specific functionality is being added.

3.1.6 Spatial Relations (Depth Area Duration Study).

Depth Area Duration (DAD) reductions for areas from 10 to 400 square miles are being updated for the entire United States and will be presented in a report separate from NOAA Atlas 14. (See previous progress report and section 4.4)

4. Progress in this Reporting Period.

4.1 Data Collection and Quality Control

HDSC updated the bulk of the daily dataset (1949-2000) with current NCDC data. With this update, we benefitted from the improved quality of daily data from NCDC and the application of our new QC methods. The quality control of extreme precipitation values for the entire Semiarid dataset is now complete.

During the initial L-moment analysis, it became apparent that the annual maximum series (AMS) of some stations had 20 or more years of values missing. Large gaps (i.e., sequential missing years) in an AMS cause concern over the data series consistency. It is not possible to guarantee that two given data segments are from the same population (same climatology, same rain gauge, same exact physical environment) from one side of the gap to the other. The causes of the missing data were unknown. Special attention was paid on stations with large gaps.

All data records were screened for large gaps using specified “Gap Check” criteria before the data was used in L-moment analysis. Software was developed for this purpose. Station records with large gaps were flagged and examined on a case by case basis using a conservative approach. Nearby stations were inspected for concurrent data years to fill in the gap if they passed the statistical test for consistency. Latitude, longitude and elevation were taken into account when examining nearby stations. Also, if there were sufficient number of years (at least 10 years of data) in each data segment, a t-test was conducted on the two segments to assess the statistical integrity of the data record. To produce more congruent data records for analysis station record lengths were adjusted where appropriate .

Finally, we developed and tested software to extract annual maximums of longer duration events based on the annual maximum criteria. Annual maximum series for 4-day through 60-day precipitation accumulations were extracted.

4.2 L-moment Analysis and Regionalization

L-moment analysis is complete for n-minute and hourly data. Initial L-moment analyses of the daily and 24-hour annual maximum series have been conducted for all 24 regions. Stations in which the discordancy value is equal to or greater than 3.0 are scrutinized for suspicious or unusual data. Methods to examine suspicious data include, but are not limited to, comparing with other data sources (e.g., archived NOAA Climatological Data), comparing with the data of other stations in the vicinity.

Based on current results, the Generalized Extreme Value (GEV) distribution is in general the best distribution for the Semiarid study area. Of the original 24 regions, 18 were found to be heterogeneous. Possible reasons for this increase in heterogeneity include:

1. switching from partial duration series to annual maximum series introduced higher L-CV especially for a semiarid area where many stations may register zero rainfall for an entire year,
2. longer record lengths introduced higher L-CV (coefficient of variation),
3. there is a greater possibility for trends or shifts in data since adding the pre-1949 dataset (TD3206).

We are currently subdividing the heterogeneous regions into smaller homogeneous regions. The process is arduous and requires explicit examination of station data and statistics with regard to climatology and topography. During this process cases where the maximum observed value at a station exceeds the 100-yr return frequency estimate are flagged for the "real data check." Effort is made during the subdivision process to mitigate such discrepancies that could be caused by (1) sample error due to small sample sizes, or (2) regionalization that does not reflect local situation. Flagged stations may be candidates for at-site analysis. However, if the at-site analysis yields precipitation frequency estimates sufficiently similar to the regional analysis, the station will be kept in the region despite a high heterogeneity value unless there is an obvious local physical reason to support its exclusion.

Table 3 shows the current subdivision progress. Completed subdivisions are pending the results of surrounding regions and an examination of mapped quantiles. The heterogeneity measure, H1, is shown for regions that still require subdivision. H1 is based on L-coefficient-of-variation (L-CV) as described in Hosking and Wallis (1997). A threshold of 2 is reasonable for the heterogeneity test, especially for precipitation data. Therefore, a value greater than 2 ($H1>2$) indicates heterogeneity, rather than homogeneity ($H1<2$).

Table 3: Initial L-moment results and subsequent subdivision progress.

Region	Requires subdivision	Results of subdivision	Status
1	yes	(H1 = 2.58)	
2	no	no subdivision required	done
3	yes	(H1 = 5.52)	
4	yes	(H1 = 5.29)	
5	no	no subdivision required	done
6	yes	4 subregions	done
7	yes	4 subregions	done
8	yes	3 subregions	done
9	no	no subdivision required	done
10	yes	4 subregions	done
11	yes	3 subregions	done
12	yes	4 subregions	done
13	yes	3 subregions	done
14	yes	5 subregions	done
15	yes	3 subregions	done
16	no	no subdivision required	done
17	yes	2 subregions	done
18	yes	(H1 = 2.23)	
19	yes	(H1 = 2.27)	
20	yes	(H1 = 6.64)	
21	yes	(H1 = 5.90)	
22	no	no subdivision required	done
23	yes	(H1 = 2.21)	
24	no	no subdivision required	done

Some of the new cross-boundary RGF differences for the sub-regions are large (i.e. cliffs). Such RGF "cliffs" are possibly the result of:

1. High skewness associated with annual maximum series at stations in a region,
2. Desert climate,
3. Statistical difficulties associated with highly variable precipitation data and zero or near-zero annual maximum values.

In particular, regions 10, 11, and 12 exhibit RGF "cliffs." If the subregions satisfy all of statistical tests, the resulting quantiles are what we expect and we've exhausted other subdivision schemes, spatial smoothing the RGF's across boundaries using appropriate criteria to determine the width of the smoothed area may mitigate this "cliff" problem.

4.3 Precipitation Frequency Data Server

Development is underway to add functionality to the Precipitation Frequency Data Server (PFDS) to extract station-specific data. Until now, users could only select a longitude/latitude location or an area, but soon the PFDS will have a pull-down menu to select a specific climate station. The menu of climate stations will represent the same stations used in the study, including the option of choosing which type of gage data (N-minute, hourly, or daily) to extract. Likewise, the data will be the exact data as output by the L-moment software used in the study. This functionality will allow for a review of the point-precipitation frequency estimates before the interpolated grids are finalized. Currently, due to co-located hourly and daily precipitation frequency estimates, a clickable station map is not available. A station map will be created to show the names and locations of the stations that will be selectable from the pull-down menu. When final estimates have been established at each station, a clickable station map will be established.

4.4 Spatial Relations (Depth Area Duration Study)

Depth Area Duration (DAD) reductions for areas from 10 to 400 square miles are being updated for the entire United States and will be presented in a report separate from NOAA Atlas 14. This quarter the focus has been on gathering and formatting data from geographically spaced dense area rain gage networks (DRNs) across the United States. These DRNs will be used in conjunction with NCDC hourly stations to develop DAD relationships. Thirteen networks have been identified thus far and are summarized in the table below.

Table 4. Dense Area Rain Gage Networks.

DRN	Period of Record (Concurrent)	Number of Stations
Coshocton, OH	1940 - 1990	10
Riesel, TX	1968 - 2001	21
Walnut Gulch, AZ	1955 - 1990	18
Reynolds Creek, ID	1965 - 1996	52
Tifton, GA	1968 - 1980	45
Alamogordo Creek, NM	1955-1977	66
Hastings, NE	1939-1962	10
Safford, AZ	1939-1971	11
Hawaii (NCDC data)	1965-2000	32
Danville, VT	1960-1974	13
Blacksburg, VA	1957-1972	15
Goodwin, MI	1981-2001	67
Lafayette, IN	1940-1953	8

5. Issues.

5.1 Updating Precipitation Frequency Atlases for Entire Nation

HDSC is currently updating the precipitation frequency atlases for a number of areas across the country and has been asked to expand the work to the entire country. Studies are underway for the Ohio River Basin and surrounding states, the Semiarid Southwest, Hawaii, and Puerto Rico and the Virgin Islands. Quarterly progress reports, which include schedules, for these studies are available at http://www.nws.noaa.gov/oh/hdsc.

Precipitation frequency studies are performed using funds provided by other federal, state and local agencies. HDSC is participating in an effort to assemble funds to update the precipitation frequency atlases for the entire United States. Hopefully sufficient funds can be identified to begin work during the summer of 2002. The full national update will use a consistent technical approach to data preparation, frequency analysis and mapping, as well as a consistent and more user-oriented approach to publication.

6. Projected Schedule.

The following list provides a tentative schedule with completion dates. Brief descriptions of tasks being worked on next quarter are also included in this section.

Data Collection and Quality Control [complete]
L-Moment Analysis/Frequency Distribution [April 2002]
Spatial Interpolation [October 2002]
Precipitation Frequency Maps [November 2002]
Temporal Distributions of Extreme Rainfall [complete]
Trend Analysis [May 2002]
Seasonal Analysis [May 2002]
Implement Precipitation Frequency Data Server [November 2002]
Write hard copy of Final Report [December 2002]
Implement review by peers [May 2002]
Publish hard copy of Final Report [February 2003]

Spatial Relations (Depth Area Duration Studies) [January 2003]

6.1 Data Collection and Quality Control.

Quality control is an iterative process that continues throughout the process. Threshold checks are complete for all data. Discordancy checks will be completed as the L-moment analysis proceeds.

6.2 L-Moment Analysis/Frequency Distribution.

Initial L-moment statistical analyses are complete for all daily, hourly and n-minute data. GEV is the best-fit distribution for the entire Semiarid study area. Careful consideration is being given to the subdivision of heterogeneous regions, real data that exceeds the 100-yr estimate, and large RGF differences between regions.

6.3 Trend Analysis and Seasonal Analysis.

The completed dataset will be analyzed during the next month for any trends or shifts in annual maximums through time. This task will involve running existing statistical software and analyzing results. A seasonal analysis will be conducted comparing the percentage of precipitation maximums that occur in each month of a year. The end products of these tasks are analyses and graphs that will be included in the final document.

6.4 Temporal Distributions of Extreme Rainfall.

The methodology we used for developing temporal distributions of extreme rainfall events will be further researched and we will verify our results. Our method is based on an Illinois State Water Survey Report (Huff, 1990) and determines the maximum and median precipitation event time distributions for 12, 24 and 72 hour duration events. Time distributions of hourly maximum and median events are sorted, averaged and plotted by storm area, quartile, duration and season.

6.5 Spatial Interpolation.

Due to contracting and data preparation delays, the period of performance for the PRISM gridding contract has been adjusted to April 1, 2002 through December 31, 2002. A kickoff conference call will initiate the task in April; implementation of project schedules and tasks will be discussed. A status meeting is tentatively scheduled in Silver Spring in July or August to discuss the interpolation methodology and draft maps.

6.6 Precipitation Frequency Data Server

PFDS changes planned for next quarter include extending the return period to 1000 years and changing the precipitation frequency estimate graph from a bar to a line graph.

6.7 Spatial Relations (Depth Area Duration Study)

Research into selecting the method to be used for computing the DAD curves will continue. Software to decode and format the data files and the DAD computations will continue to be developed. As DRN's are located, they will be added to our database.

References

Frederick, R.H., V.A. Myers and E.P. Auciello, 1977: Five- to 60-minute precipitation frequency for the eastern and central United States, NOAA Technical Memo. NWS HYDRO-35, Silver Spring, MD, 36 pp.

Hershfield, D.M., 1961: Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, *Weather Bureau Technical Paper No. 40*, U.S. Weather Bureau. Washington, D.C., 115 pp.

Hosking, J.R.M. and J.R. Wallis, 1997: *Regional frequency analysis, an approach based on L-moments*, Cambridge University Press, 224 pp.

Huff, F. A., 1990: Time Distributions of Heavy Rainstorms in Illinois. Illinois State Water Survey, Champaign, 173, 17pp.

Lin, B. and L.T. Julian, 2001: Trend and shift statistics on annual maximum precipitation in the Ohio River Basin over the last century. Symposium on Precipitation Extremes: Prediction, Impacts, and Responses, 81st AMS annual meeting. Albuquerque, New Mexico.

Miller, J.F., 1964: Two- to ten-day precipitation for return periods of 2 to 100 years in the contiguous United States, *Technical Paper No. 49*, U.S. Weather Bureau and U.S. Department of Agriculture, 29 pp.

Miller, J.F., R.H. Frederick and R.J. Tracy, 1973: Precipitation-frequency atlas of the western United States, *NOAA Atlas 2*, 11 vols., National Weather Service, Silver Spring, MD.